

AD-A073 365

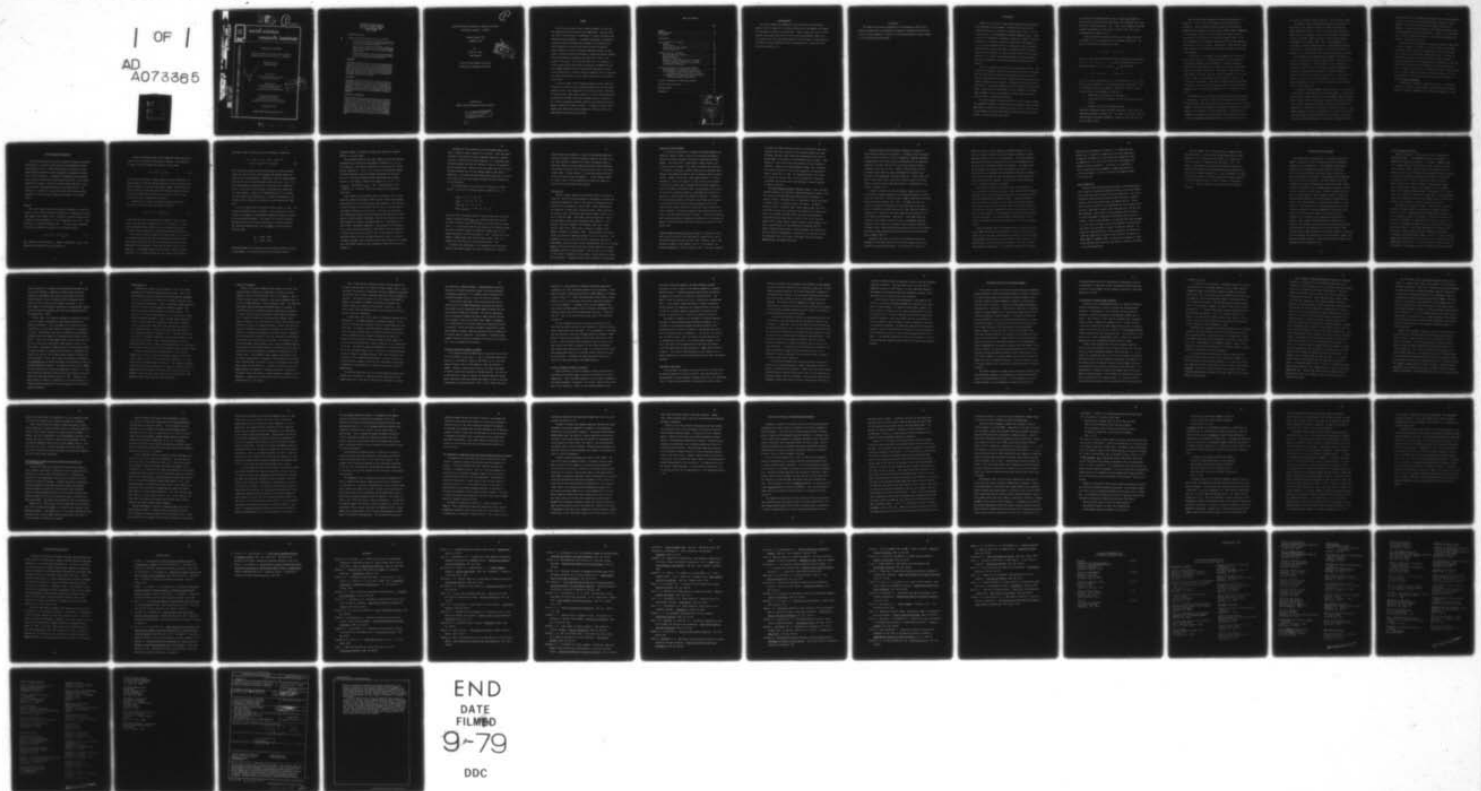
DECISIONS AND DESIGNS INC MCLEAN VA  
IMPORTANCE WEIGHT ASSESSMENT FOR ADDITIVE, RISKLESS PREFERENCE --ETC(U)  
DEC 78 R S JOHN, W EDWARDS

F/G 5/10  
N00014-76-C-0074

UNCLASSIFIED

NL

| OF |  
AD  
A073365





RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL II

①

001922-5-T

AD A073365

DDC FILE COPY



USC

UNIVERSITY OF SOUTHERN CALIFORNIA

# social science research institute

## RESEARCH REPORT

IMPORTANCE WEIGHT ASSESSMENT FOR ADDITIVE,  
RISKLESS PREFERENCE FUNCTIONS: A REVIEW

RICHARD S. JOHN  
WARD EDWARDS

### SPONSORED BY:

ADVANCED RESEARCH PROJECTS AGENCY  
DEPARTMENT OF DEFENSE

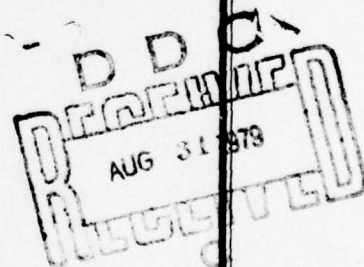
### MONITORED BY:

ENGINEERING PSYCHOLOGY PROGRAMS  
OFFICE OF NAVAL RESEARCH  
CONTRACT No. N00014-76-C-0074, ARPA

APPROVED FOR PUBLIC RELEASE;  
DISTRIBUTION UNLIMITED;  
REPRODUCTION IN WHOLE OR IN PART IS  
PERMITTED FOR ANY USE OF THE U.S. GOVERNMENT

DECEMBER 1978

SSRI RESEARCH REPORT 78-5



79 08 31 059

B

**Social Science Research Institute  
University of Southern California  
Los Angeles, California 90007  
(213) 741-6955**

**INSTITUTE GOALS:**

The goals of the Social Science Research Institute are threefold:

- To provide an environment in which scientists may pursue their own interests in some blend of basic and methodological research in the investigation of major social problems.
- To provide an environment in which graduate students may receive training in research theory, design and methodology through active participation with senior researchers in ongoing research projects.
- To disseminate information to relevant public and social agencies in order to provide decision makers with the tools and ideas necessary to the formulation of public social policy.

**HISTORY:**

The Social Science Research Institute, University of Southern California, was established in 1972, with a staff of six. In fiscal year 1978-79, it had a staff of over 90 full- and part-time researchers and support personnel. SSRI draws upon most University academic Departments and Schools to make up its research staff, e.g. Industrial and Systems Engineering, the Law School, Psychology, Public Administration, Safety and Systems Management, and others. Senior researchers have point appointments and most actively combine research with teaching.

**FUNDING:**

SSRI Reports directly to the Executive Vice President of USC. It is provided with modest annual basic support for administration, operations, and program development. The major sources of funding support are federal, state, and local funding agencies and private, foundations and organizations. The list of sponsors has recently expanded to include governments outside the United States. Total funding has increased from approximately \$150,000 in 1972 to almost \$3,000,000 in the fiscal year 1978-1979.

**RESEARCH INTERESTS:**

Each senior SSRI scientist is encouraged to pursue his or her own research interests, subject to availability of funding. These interests are diverse: a recent count identified 27. Four major interests persist among groups of SSRI researchers: crime control and criminal justice, methods of dispute resolution and alternatives to the courts, use of administration records for demographic and other research purposes, and exploitation of applications of decision analysis to public decision making and program evaluation. But many SSRI projects do not fall into these categories. Most projects combine the skills of several scientists, often from different disciplines. As SSRI research personnel change, its interests will change also.



(P)

Importance Weight Assessment for Additive, Riskless  
Preference Functions: A Review

Research Report 78-5

December, 1978

by

Richard S. John

Ward Edwards



Social Science Research Institute  
University of Southern California

Sponsored by

Defense Advanced Research Projects Agency

This document has been approved  
for public release and sale; its  
distribution is unlimited.

79 08 31 059

## SUMMARY

One of the more useful tools in decision analysis is the riskless, additive multi-attribute utility (MAU) model. The most difficult task in the application of MAU models is that of estimating the importance weight parameters. Two general approaches to the weight estimation problem are extensively reviewed in the present paper: direct subjective estimation and indirect holistic estimation. Various methods for directly assessing importance weights are catalogued, including ranking, fractionation, subjective-estimate methods, and paired-comparison procedures, and their relationship to one another is discussed. The so-called indirect holistic methods, including unbiased and biased regression analyses, the ANOVA and fractional ANOVA paradigms, and the indifference techniques of pricing out and trading off to the most important dimension, are all explained with particular emphasis on their common relationship to the general linear model.

A critical review of the literature comparing direct subjective estimation to indirect holistic methods revealed that the conclusions reached by Slovic and Lichtenstein in 1971 are no longer justifiable, if they ever were. Many recent studies are cited in which subjective and statistical weights yield high convergent validity, contrary to the "serious discrepancies between subjective and objective relative weights" focused on in earlier reviews. In addition, several recent studies have established strong evidence of criterion validity for both subjective and statistical weights.

# TABLE OF CONTENTS

	Page
Summary .....	i
Acknowledgements .....	iii
Disclaimer .....	iv
Introduction .....	1
Direct Subjective Estimation .....	6
Ranking .....	6
Fractionation .....	11
Subjective Estimate Methods .....	12
Paired Comparisons .....	16
Indirect Holistic Estimation .....	18
Multiple Regression Analysis .....	19
Ridge Regression .....	21
Analysis of Variance .....	22
Fractional Replication Designs for ANOVA .....	24
Holistic Orthogonal Parameter Estimation .....	25
Indifference Techniques .....	26
Convergent Validation of Importance Weights .....	29
Correspondence of Actual Weight Estimates .....	30
Correspondence of Composites Derived from Assessed Weights and Holistic Evaluation.....	34
Correspondence of Composites Resulting from Different Weighting Procedures .....	38
Criterion Validation of Importance Weights .....	41
Conclusion and Discussion .....	48
Reference Notes .....	49
References .....	51

Accession For	
NTIS Grant	<input checked="checked" type="checkbox"/>
DDC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution	
Availability Codes	
Dist	Avail and/or special
A	

#### ACKNOWLEDGMENTS

This project benefitted immeasurably from discussions with Hut Barron, Norman Cliff, Bill Gabrielli, Jim Kahan, Norman Miller, Bob Newman, Dave Seaver, Bill Stillwell, and Detlof von Winterfeldt. Special thanks are given to Jenee Euchner, who provided both technical editing and encouragement. Support for this research came from the Advanced Research Projects Agency of the Department of Defense, contract N00014-76-C-0074, on subcontract P.O. 78-072-0720 from Decisions and Designs, Inc.



#### DISCLAIMER

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or of the United States Government.



## Introduction

Research in the field of human choice behavior has traditionally been segregated into two domains: description and prescription. Although the two have conceptually different goals, there is a striking similarity in the mathematical problem statement of each. Given a set ( $S$ ) of alternatives, strategies, objects, or courses of action, the problem is to define a function ( $f$ ) which maps the elements of  $S$  into some well-ordered set that is isomorphic to the real numbers. Both begin with a set  $S$ , both seek to capture the "true" preference structure of an individual or group, and both finally arrive at some mapping  $f$ . In the end, each approach produces a mapping intended to abstract key components of cognitive mechanisms determining choice behavior.

The task of specifying  $f$  presents two interrelated problems: the nature of the judgments required, and the functional form of the model. Any functional form will usually impose some restriction on the elicitation procedures, and vice versa. Most prescriptive and descriptive research compares, tests, explores, or otherwise validates particular model forms in some specific domain of choice alternatives and decision-makers. Most of what is known about eliciting judgments to estimate model parameters comes from psychophysics rather than psychometrics, decision theory, or cognitive psychology.

This paper compares various methods of eliciting responses needed for parameter estimation for additive choice structures under certainty. Most choice situations are non-additive, risky, or both. But more complicated models, involving more complex function forms and more parameters, are often not worth the effort.

In particular, the additive model serves as a good approximation to much more complicated function forms (Dawes, 1971; Goldberg, 1965, 1968, 1970, 1971; Yntema & Torgerson, 1961). Also, risky preference functions are well approximated by riskless ones (Fischer, 1976, 1977, Note 1; von Winterfeldt & Edwards, Note 2).

Under assumptions of riskless additivity, the composite worth of any element of  $\underline{S}$  is a weighted sum of its component attributes. The problem of specifying a mapping  $f$  such that

$$C_i = f(X_i) \quad (1 \leq i \leq n) \quad (1)$$

reduces to constructing  $k$ -dimensional vectors representing each element of  $\underline{S}$ , i.e.,  $X_i = (x_{i1}, x_{i2}, \dots, x_{ik})$ , and a  $k$ -dimensional vector of importance weights, i.e.,  $W = (w_1, w_2, \dots, w_k)$ , such that

$$C_i = w_1 x_{i1} + w_2 x_{i2} + \dots + w_k x_{ik}, \quad (1 \leq i \leq n). \quad (2)$$

( $C_i$  is the composite evaluation of  $X_i$ , the  $i$ th member of  $\underline{S}$ ;  $x_{ij}$  is the value of the  $i$ th member of  $\underline{S}$  on the  $j$ th attribute or dimension.) Once  $\underline{S}$  is defined and the decision-maker is identified, three main steps serve to specify the parameters in Equation 2:

1. Determine the  $k$  dimensions of importance;
2. Locate each of the  $n$  elements of  $\underline{S}$  in a  $k$ -dimensional vector space;
3. Generate a vector of importance weights.

The various approaches impose different orderings of these steps and some even eliminate or combine them. This paper is primarily concerned with estimating the weight parameters, though the other two steps are also in need of study.

There are two general approaches to obtaining the W vector in Equation 2: direct subjective estimation and indirect holistic estimation. Direct subjective weight estimates are derived from judgments about abstractions of the choice problem, namely, dimensions. It is important to keep in mind that dimensions (whether directly measurable or not) exist only in the decision-maker's head. An explicit awareness of the relative importance of the various dimensions identified in step 1 is assumed. Operationally, direct methods (sometimes referred to as decomposition techniques) typically require only a few judgments.

Indirect holistic procedures require judgments that directly relate to some subset of S. Magnitude estimates (or the equivalent) of particular elements of S are treated as "criterion" measures; weights are derived indirectly via statistical estimation procedures based on the general linear model. (The general linear model is so named because of assumed linearity in weight parameters, not in scale values.) There is no assumption that the subject is aware of the "dimensions of importance," upon which the alternatives are evaluated. In practice, indirect holistic techniques usually require many more judgments than the direct ones; however, this is not necessarily the case.

Much attention has focused on the prospect that weight assessment is unnecessary. Since Wilks (1938) first published on the robustness of equal weights, many have argued that the weighting question is trivial. Indeed, in the area of psychometrics, differential weighting of component scores of a test battery is all but nonexistent. The mental tests literature is replete with formal analytic work demonstrating the excellent correspondence between different sets of composites derived

from different weighting schemes, (Ghiselli, 1964; Gulliksen, 1950).

This wheel has been rediscovered many times, most recently in the areas of human judgments and decision-making and multiple linear regression analysis. There is now little doubt that when dimensions are uncorrelated or positively correlated, any weighting scheme is acceptable (Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975; Newman, 1977; Wainer, 1976, 1978). Given agreeable (i.e., non-negative) intercorrelation matrices, it hardly matters whether the weights are obtained subjectively, statistically, randomly, or a priori (i.e., equal weights); the results are essentially the same.

Recent arguments from both the human judgment and regression analysis literatures have strongly challenged the "non-negative inter-correlation assumption." Calling attention to the importance of suppressor variables in multiple regression, Keren and Newman (1978) rejected the equal-weighting approach as a general methodology. Negative correlations are present, by definition, in the case of suppressor variables; thus, the one assumption critical to the unit weighting argument is simply not met in at least this one important case of linear regression.

In the area of human judgment and decision-making, the assumption of non-negative intercorrelations is even more tenuous. Drawing upon previous work by Edwards and his associates (Newman, Seaver, & Edwards, Note 3; Seaver, Note 4), McClelland (in press) proved that attributes will be highly negatively correlated if the domain of alternatives (the set S) is restricted to only those on the Pareto frontier. (The Pareto frontier of any set of alternatives, S, consists of those that are not dominated. Although dominance may be defined in many ways, an ordinary dominated alternative is one that is no better than some



other alternative on all dimensions and worse on at least one dimension.) Of course, for the task of either describing or prescribing choice behavior, only those alternatives on the Pareto frontier are of interest. By adding various dominated (irrelevant) alternatives, one could generate any intercorrelation matrix. However, if an alternative has no chance of being chosen (which is the case for dominated alternatives), why consider it at all?

Working only with alternatives on the Pareto frontier, McClelland (in press), Newman et al. (Note 3), and Seaver (Note 4), concluded that composites derived from unit weighting will not agree satisfactorily with those obtained from differential weights. In addition, McClelland (in press) showed that the overall value of the best composite determined from unit weighting may be substantially less than that obtained from the correct differential weights, where overall value is computed using the "true" differential weights. These analytic results strongly suggest that the equal weighting argument is simply not applicable to the multi-attribute problem in decision-making. In short, the weights do matter.

An anthology of approaches to weighting follows. It describes and gives a rationale for each method. When appropriate, a detailed account of necessary judgmental and arithmetical procedures is given.



### Direct Subjective Estimation

The task of any direct subjective estimation strategy for defining the W vector is to create a ratio scale for the importance of dimensions defined in step 1. Most direct estimation procedures use well-known techniques prominent in the psychophysics and general psychological scaling literature. The transfer of methodology from psychological scaling to the problem of specifying the W vector is not without complication, however. Unlike the concrete stimuli employed in most scaling studies, importance dimensions are but abstractions. The following is a discussion of some of the more widely used assessment procedures for obtaining direct subjective estimates of importance weights.

#### Ranking

One of the simplest subjective estimation procedures for obtaining weights is that of rank-ordering the dimensions in importance (Eckenrode, 1965; Newman, 1977; Permut, 1973). Typically, the decision-maker places a numeral beside each member of a list of dimensions, such that 1 = most important, 2 = next most important, etc. Ties are usually permitted. The ranks are converted to weights by the following formula:

$$W_i = (k + 1 - R_i) / \sum_{j=1}^k R_j \quad (3)$$

(W<sub>*i*</sub> = weights on *i*th attribute, k = number of dimensions, and R<sub>*i*</sub> = subjective rank assigned to the *i*th alternative.)

Stillwell and Edwards (Note 5) have suggested alternative ways to convert the rank-ordering of attributes to weights. One suggestion is that of using the normalized reciprocals of the ranks:

$$W_i = (1/R_i) / \sum_{j=1}^k 1/R_j \quad (4)$$

To cope with the problem of choosing among the three alternatives of equal weighting, "rank sum" weights (Equation 3), and "rank reciprocal" weights (Equation 4), Stillwell and Edwards suggest that the decision-maker be asked to provide the weight for the most important dimension and that the method which delivers a value for  $W_1$  closest to that estimated be used in determining the  $W$  vector.

Alternatively, Stillwell and Edwards propose another transformation of the ranks into "rank exponent" weights:

$$W_i = (k + 1 - R_i)^z / \sum_{j=1}^k (R_j)^z \quad (5)$$

To determine the value of the arbitrary constant  $z$ , one elicits  $W_1$  subjectively and solves iteratively for  $z$ . (In practice,  $z$  is most easily determined from a table set up as a function of  $W_1$  and  $k$ .) The entire vector of weights may then be determined from Equation 5.

Using results derived by Abelson and Tukey (1963), it is possible to transform ranks to weights which are certain to provide the maximum, minimum correlation with the true weights given that the rank-ordering of the true weights is that same as that of the elicited ranks,  $R_i$ . The general formula for the "maximim" weights, given in Equation 6, is a slightly modified version of that given by Abelson

and Tukey (1963), allowing the sum of the weights to equal one.

$$W_i = \{ [ (R_i - 1) (1 - ((R_i - 1)/k)) ]^{.5} - (R_i (1 - R_i/k))^{.5} \} + A / (k \cdot A). \quad (6)$$

Since the maximin weights are only defined up to a linear transformation, any choice of  $A$  will yield weights with maximum minimum correlation with the true weights. Ordinarily,  $A$  will be chosen greater than one to yield all positive weights that sum to one. For  $A = 1.0$ , the spread among the weights will be close to the most extreme possible. As the choice of  $A$  increases, the weights will become more like equal weights. Clearly, the indeterminacy of  $A$  is a problem. One could obtain many different sets of weights from Equation 6, all of which satisfy the maximin criterion, by simply varying the choice of  $A$ .

One brief comment about the maximin criterion seems in order. The correlation between two sets of weights is only indirectly (at best) related to the degree of correspondence between the composites defined by the two weighting schemes. For example, the two sets of weights defined below are, for all practical purposes, "extreme" weighting ( $W_1$ ) and equal weighting ( $W_2$ ), yet the weights correlate perfectly with each other.

$$W_1 = (.9999, .0001)$$

$$W_2 = (.5001, .4999)$$

The appropriateness of the maximin criterion between weighting schemes is questionable. A criterion that does not distinguish between

weighting schemes so different as about equal weights and extreme weights is not very useful.

Also reported by Abelson and Tukey (1963) are the actual maximin correlations as a function of the number of dimensions,  $2 \leq k \leq 20$ . As a comparison, the minimum correlations are also given for "linear" weights (Equation 3). For  $k = 2$  or  $3$ , both of the weighting schemes correlated perfectly; thus the minimum correlations are the same. The loss in correlation between weights does become quite severe, however, as the number of dimensions increases (e.g., for  $k = 20$ ,  $r_{\text{maximin}} = .64$ , whereas  $r_{\text{linear}} = .38$ ). The usefulness of such a comparison is, of course, called into question by the arguments given above.

Other schemes for transforming ranks into weights, such as that proposed by Mosteller and Tukey (1977), suffer from the same indeterminacy problems as the maximin weights described above. The Mosteller and Tukey procedure yields weights that sum to zero, and thus must be transformed (linearly?), to sum to one. Of course, one could cure the indeterminacy problem for both the maximin and Mosteller and Tukey weights by obtaining more information from the decision-maker, as has been proposed by Stillwell and Edwards (Note 5). The usefulness of the Stillwell and Edwards suggestion is questionable for two reasons: (1) a single subjective estimate of "the weight of the most important attribute" may not be valid, and (2) even if the weight for the most important dimension can be estimated accurately, there is no reason to expect, a priori, that a weighting scheme that reproduces this weight will yield better composites than a weighting scheme that does not reproduce it.



Although the first problem is yet to be answered empirically, there is reason to remain skeptical at this point. Since the number elicited as the weight on the most important dimension is meaningful only in relation to the other dimensions, it is imperative that the decision-maker consider the importance of all of the dimensions when he estimates  $W_1$ . Asking for only one number runs the risks that the decision-maker will not give enough thought to the other  $k - 1$  dimensions; If he does think hard enough about them, there is little justification for not going ahead and obtaining direct estimates of all of the weights.

The second problem is more directly illustrated by a simple example. Consider the following weighting schemes (for  $k = 4$ ):

$$W_{\text{true}} = (.4, .22, .20, .18)$$

$$W_{\text{equal}} = (.25, .25, .25, .25)$$

$$W_{\text{rank sum}} = (.4, .3, .2, .1)$$

$$W_{\text{rank reciprocal}} = (.48, .24, .16, .12)$$

By the Stillwell and Edwards method of choosing among the last three schemes, one would choose  $W_{\text{rank sum}}$ , given that the decision-maker was able to correctly specify the weight on the most important dimension, .4. If one assumes that the attributes are not intercorrelated, then the correlation between the composites formed from the true weights and those from rank-sum weights is .978. Although this correlation is indeed higher than that for equal weights, .944, it is slightly less than that for rank-reciprocal weights, .983.

Of course, the differences in this example are not very great; however, the example suggests the type of problem one is likely to



encounter when such criteria as "choose the weighting scheme that matches the weight on the most important dimension" are used (e.g., the "rank exponent" procedure of Stillwell and Edwards, Note 5). With more dimensions and non-negative attribute intercorrelations, the problems are quite likely to be more serious than in the simple example shown. Further research is clearly warranted on the topic of how to transform ranks into weights. As the situation stands at present, there is little theoretical or empirical reason for preferring one transformation over another.

#### Fractionation

The most commonly employed fractionation method (as defined by Torgerson, 1958) for assessing weights is the method of constant sum, advocated by Mettessel (1947) in the context of psychophysics and formalized more completely by Comrey (1950). After rank-ordering the dimensions of importance, the decision-maker, under instruction to preserve ratios, distributes some constant number of points (e.g., 100 or 1000) over them. The weight on the  $i$ th dimension is simply the percentage of total points assigned to that dimension. (For examples, see Cook & Stewart, 1975; Hoffman, 1960; Klahr, 1969; Schmitt, 1978; Slovic, 1969; Slovic, Fleissner, & Bauman, 1972; Summers, Taliaferro, & Fletcher, 1970). The advantage of this methodology for assessing importance weights lies in its simplicity. The judgments are only a trifle more difficult than those required for rank weights, and the troubles of rank transformations are easily avoided. However, distributing points necessarily focuses attention on weight differences and not on weight ratios. Since the ratio of the weights is the critical information, serious biases in fractionation estimates may be prevalent. No empirical test of this hypothesis has been made.

### Subjective Estimate Methods

The two-way classification of subjective-estimate methods suggested by Torgerson (1958) is useful for discussing the numerous subjective-estimate procedures utilized for obtaining importance weights. One distinction is made between "single stimulus" and "multiple stimuli" methods. This refers to whether each dimension is followed by a response or whether dimensions are presented simultaneously and numbers assigned to each in whatever order the subject decides. Another distinction is made between methods of "limited categories," in which the subject is given a finite set of numbers into which the stimuli must be mapped, and "unlimited categories," wherein the numbers assigned to stimuli are generated by the subject. In the case of unlimited categories, respondents may be asked for graphical or spatial rather than numerical responses. Virtually all of the importance weighting techniques, except those using ratings, fit into the multiple-stimuli/unlimited-categories cell of the classification scheme; ratings are a multiple-stimuli/limited-categories procedure. We found no examples of any single-stimulus technique in the importance weighting literature. The primary differences among subjective-estimate procedures for weights are in the number and nature of the anchor points that specify the origin and unit for the weight scale.

Limited categories methods (multiple stimuli). An approach closely akin to that of ranking is the technique of categorizing or rating. The decision-maker puts each attribute into a category, usually identified by a numeral, often between 1 and 10. (For examples, see Nystedt and Magnusson, 1975, and Schmitt, 1978.) The usual procedure

for obtaining weights from the ratings is to perform the same transformation specified in Equation 3 for obtaining weights from ranks. Of course, the same issues concerning transformation from ranks to weights apply to the problem of transforming rates to weights. Rate weighting is even simpler than rank weighting for small numbers of categories. However, the rating approach produces less information than ranking, especially if only a few categories are used and the number of dimensions is large. For some applications, this insensitivity may be a virtue, but it limits the amount of information a respondent can provide, and so may encourage a careless approach to response selection.

Unlimited categories methods (multiple stimuli). Most unlimited-categories methods are versions of magnitude estimation (Stevens, 1957). The primary distinction among magnitude-estimation procedures is whether a modulus (reference point) is presented to the decision-maker or whether he/she is allowed to choose the modulus. No-modulus examples include: Eckenrode (1965), whose subjects drew lines from attributes to points on a continuous line marked off in integer units from 0 to 10; Hoepfl and Huber (1970), who used the same method, except that the scale went from 0 to 100 in units of 10; Cook and Stewart (1975), whose subjects simply assigned a number between 0 and 100 to each attribute; and von Winterfeldt and Edwards (Note 2), wherein subjects placed slashes on nine-centimeter lines without endpoints or numerical segmentation. Weights were the distance from the "not important" end of the lines or the number directly estimated, depending upon the response mode used.

Importance weights have often been estimated in relation to some reference dimension of importance. For example, Cook and Stewart (1975) assigned a weight of 100 to a "moderately" important dimension, as determined by the subject, and elicited weights for the remaining dimensions such that "the ratio of the ratings reflected relative importance of the cues (p. 35)." Schmitz (1978) utilized a similar procedure, except that the subject determined the constant to be assigned to the "moderately" important dimension. Fischer (1976, 1977, Note 1) and O'Connor (1972) have used a slight variation on this approach, assigning the most important dimension, as determined by the subject, a weight of 100.

A number of studies have been performed by Edwards and his associates within a general methodology for determining additive, riskless multi-attribute utility functions known as SMART (Simple Multi-Attribute Rating Technique) (Edwards, 1972, 1977). The SMART procedure prescribes that weights be estimated by first rank-ordering the dimensions of importance, and then assigning the least important dimension a weight of 10. Weights on the other dimensions are assessed in the ratio fashion outlined above. Applications of the SMART weighting technique can be found in the area of land use management (Gardiner & Edwards, 1975), planning a government research program (Guttentag & Snapper, 1974), credit-card applicant evaluation (Eils & John, Note 6), and choosing among alternative bussing plans for court-ordered segregation (Edwards, Note 7).

Eckenrode (1965) used a variation of the magnitude-estimation procedure to facilitate checking for consistency among weight ratios implied by the magnitude estimates. In the so-called "successive



comparison" method, subjects began by assigning a weight of 1.0 to the most important attribute, and distributing the weights for other attributes between 0 and 1, as discussed above. Next, each dimension, beginning with the most important, was successively compared to the set of all other dimensions ranked less important. The subject then decided whether the dimension under consideration was more or less important than the combination of all dimensions ranked less important. The weight of the dimension under consideration was then adjusted to be consistent with that judgment. The consistency check ended after all  $k-1$  attributes had been so evaluated.

Some limited interest has also been shown in obtaining direct ratio estimates of attribute importance. Fujii-Eustace (1978) obtained estimates of the ratio of attribute importance in a laboratory setting involving two-dimensional commodity bundles. Fischer and Peterson (Note 8) conducted a laboratory study comparing magnitude estimates of importance (with the most important dimension assigned a weight of 100) and ratio estimates (wherein the ratio of the importance of the most important dimension to the importance of each of the other dimensions was estimated). The ratio-estimation method produced significantly less uniform distributions of weights for 15 of the 16 subjects studies.

Otway and Edwards (Note 9) and Edwards (Note 7) applied the traditional SMART methodology to the problem of siting a nuclear waste disposal facility; as usual, the weights were obtained via ratio estimates. The respondents were required to judge ratios of the importance of all possible pairs of the six dimensions specified. The ratio estimates were assessed simultaneously in a triangular tableau, and consistency



among the ratio estimates was forced (i.e., if dimension A was judged twice as important as dimension B, and B was judged twice as important as dimension C, then A was either judged four times as important as C, or one or both of the first two judgments was changed to force consistency.) Of course, it is possible to think of each direct ratio judgment as a magnitude estimation in which the modulus is set to 1.0 for the less important dimension in each pair.

#### Paired Comparisons

Eckenrode (1965) employed three variations of the paired-comparisons procedure for obtaining importance weights, and Cook and Stewart (1975) used yet another variation of paired comparisons. One of the assessments administered by Eckenrode (1965) involved a triangular tableau, such as that used by Otway and Edwards (Note 9). Rather than indicating the ratio of the importances of the attributes within each pair, Eckenrode's subjects simply indicated which attribute was more important. Two variants of this procedure required subjects to circle the member of each pair of attributes which was more important. In one of the procedures, all possible pairs were presented in a list, and in the other, the list was doubled by including each pair twice, with the order of stimuli in each pair reversed. The weight for each dimension was simply calculated as the frequency of times that dimension was chosen as more important, divided by the total number of judgments made. Of course, this weight will simply be the inverted rank, given that the subject is perfectly consistent in his paired comparisons.

Cook and Stewart (1975) listed all pairs of dimensions and required subjects to indicate that the attributes were of "equal importance (0), or that one cue was slightly (1), substantially (2), or much more (3) important than the other (p. 35)." The number in parentheses was assigned to the dimension judged more important, and the weight for each dimension was calculated as the sum of all the numbers assigned to it. Such a procedure constitutes a crude approximation to the ratio-estimation procedures advocated by Otway and Edwards (Note 9), since the numbers in parentheses (0, 1, 2, 3) might be thought of as imprecise ratio estimates of importance.

### Indirect Holistic Estimation

The common defining characteristic of indirect holistic procedures to weighting is their reliance upon holistic evaluations of complex choice alternatives. Such approaches often require numerous holistic judgments and utilize specific statistical tools for analyzing covariance structures, such as multiple regression and analysis of variance. The weights are never obtained from direct subjective estimates. They are inferred within the framework of the specific mathematical (statistical) model assumed to relate the elements of  $\underline{S}$ , the unknown weight parameters, and the holistic estimates of overall worth. Holistic evaluations of elements of  $\underline{S}$  are obtained in various manners, usually via some subjective-estimate method, such as rating scales or magnitude estimation. Very little attention is given to which method is in fact employed. The primary basis for distinguishing among the various indirect approaches is the exact mathematical (statistical) model used, both to prescribe the subset of  $\underline{S}$  upon which holistic judgments must be obtained, and to mechanically determine the weights. Often, several objective measures of importance of a dimension are available, all of which are strictly equivalent when the attributes of the alternative set  $\underline{S}$  are uncorrelated. When intercorrelations among dimensions are non-zero, however, the numerous objective measures of importance will not be equivalent; indeed, they may not even agree ordinally. Discussions of the most commonly used indirect holistic approaches follow.

### Multiple Regression Analysis

Multiple regression is often used as a tool for determining importance weights. In experimental settings, this approach is theoretically dependent on Brunswik's "lens model" (Hammond, 1966) and is methodologically tied to the well-known "multiple cue probability learning" (MCPL) paradigm (Slovic and Lichtenstein, 1971). Regression analysis has also been applied extensively and with success to actual policy-capturing problems (for a review, see Slovic and Lichtenstein, 1971). While the traditional label for multiple regression applied to decision problems is bootstrapping (Dawes, 1971), a more current label for this approach is Social Judgment Theory (SJT) (Hammond, Stewart, Brehmer, & Steinmann, 1975). (There seems to be virtually no relationship between SJT in the bootstrapping context and the more well-known Social Judgment Theory of Sherif and Hovland, 1961, relating to persuasion and attitude change.)

The regression procedure is simple: obtain holistic subjective evaluations of some subset of  $\underline{S}$  and perform a standard regression analysis. (Each  $\underline{k}$  - dimensional vector, representing an element in  $\underline{S}$ , is treated as a row of the predictor " $\underline{X}$ " matrix and the entire set of holistic evaluations is treated as the criterion " $\underline{Y}$ " vector. If  $\underline{m}$  holistic assessments are made, with  $\underline{k} \leq \underline{m} \leq \underline{n}$ , then  $\underline{X}$  will be  $\underline{m}$ -by- $\underline{k}$  and  $\underline{Y}$  will be  $\underline{m}$ -by-1.) Well-known techniques and formulae may then be applied to obtain various statistical measures of importance: (1) ordinary least squares beta weights,  $\hat{\beta}$ , and  $\hat{\beta}^2$ ; (2) the correlation between the attributes and the "criterion",  $\underline{r}$  (also called the validity coefficient), and  $\underline{r}^2$ ; (3)  $\hat{\beta} \cdot \underline{r}$ ; (4) the usefulness index,  $\underline{U}$ , which is the increase in the squared multiple correlation coefficient observed



when an attribute is included with the remaining attributes of the set; and (5) Englehart's measure,  $\underline{E}$ , based on the sum of each attribute's independent effect and its joint effect with every other attribute. Any one of these statistical measures may be used as a direct estimate of the weights in Equation 2. For a good summary of the voluminous applications and experimental studies using the regression paradigm, see Slovic and Lichtenstein (1971) and Hammond et al. (1975).

The key idea in the regression approach is the notion of using statistical indices as estimates of the importance weights (for prediction purposes). If all attributes of a set have zero inter-correlations, one statistical measure of importance is the same as any other. Unfortunately, when cues or dimensions are correlated, different statistical measures of importance may be totally contradictory. This point has been demonstrated at least three times over the last sixteen years (Darlington, 1968; Schmitt & Levine, 1977; Ward, 1962). Darlington's (1968) conclusion is particularly pointed: "It would be better to concede that the notion of 'independent contribution to variance' has no meaning when predictor variables are intercorrelated (p. 166)." Which statistical measures of importance are psychologically valid when attributes are correlated is simply an unanswered empirical question. As Schmitt and Levine (1977) comment, "all the (statistical) indices are paramorphic in nature, i.e., they were derived from regression analyses which have nothing to do necessarily with the actual decision process (p. 26)." Ambiguity over which statistical measure of importance is valid when dimensions are intercorrelated substantially limits the usefulness of the multiple-regression approach.

### Ridge Regression

Ordinary least-squares (OLS) estimates of the  $\beta$ 's are highly unstable when attribute intercorrelations are high. That is, the OLS estimates may be quite different from the population parameters of the  $\beta$ 's, resulting in a substantial shrinkage in multiple correlation upon cross-validation. Since  $\beta$ 's and other related regression statistics are often used to determine importance weights, high multi-collinearity among attributes is a serious problem. One solution is to utilize biased estimates of  $\beta$ , via any one of several recently developed techniques. One of the more popular approaches is that of ridge regression, invented by Hoerl (1962). Numerous simulation studies have shown that ridge estimates are superior, in cross-validation terms, to those of the more common OLS approach (e.g., Dempster, Schatzoff, & Wermuth, 1977). Excellent discussions of the multi-collinearity problem and the solution afforded by ridge regression may be found in Hoerl and Kennard (1970a, 1970b) and Marquardt and Snee (1975). More recent presentations in a psychological context may be found in Price (1977), Darlington (1978, and Winer (1978).

As Schmitt and Levine (1977) demonstrated in a multiple-cue judgment task with high intercue correlations, weights derived from ridge estimates of the  $\beta$ 's may be completely contradictory to those derived from a standard regression analysis. The arguments presented by McClelland (in press), Newman et al., (Note 3), and Seaver (Note 4), suggesting that high negative attribute intercorrelations are to be expected in decision tasks, would lead to the conclusion that ridge weights are more useful than their OLS counterparts.

### Analysis of Variance

Analysis of variance (ANOVA) has been used as a tool for identifying relevant dimensions of importance in studies of choice behavior (see Slovic & Lichtenstein, 1971, for a summary). One of the most notable research programs in the area of judgment theory to utilize the ANOVA approach is that of Anderson (1974). The functional measurement approach developed by Anderson and his colleagues is typified by the use of factorial designs, quantitative evaluations of overall stimulus value, and monotonic transformations (Anderson, 1977). Normally, elements of  $S$  are specified as nominal levels (possibly ordered, e.g., high, medium, and low) on various descriptive dimensions (factors) of presumed importance. Usually, the subject is required to numerically (or spatially) evaluate, at least once, all elements of  $S$  created by completely crossing all factors. Depending upon the theoretical concerns of the study, monotonic transformations to eliminate interaction effects may be used. The prototypical functional measurement study proceeds to test, within the ANOVA framework, the hypothesis that the "main effects" are significant, given that all interactions may be eliminated by a monotonic transformation of the response scale. Such hypothesis tests, whether significant or not, are usually accompanied by a statement of the magnitude of the effect. Often, the magnitude of an effect is used as an estimate of the importance weight for that dimension. Virtually all analyses within the ANOVA paradigm are nomothetic, focusing upon questions concerning the manner in which the group as a whole aggregates information across dimensions. Sometimes verification that group findings hold in idiographic analyses is also given.

Since "linear multiple regression analysis and the analysis of variance (and covariance) are identical systems (Cohen, 1968, p. 426)," the statistical measures of importance derived via ANOVA are equivalent to statistics which could be generated within the multiple regression approach. Since multiple regression is more general than ANOVA (or ANCOVA), every ANOVA (or ANCOVA) problem may be formulated as a multiple regression problem, but the reverse is not true. Thus, from a formal view, the ANOVA paradigm is but a special case of the multiple regression methodology.

In application, there are a number of differences between the two methodologies. The ANOVA paradigm, with factorial designs, is not suited for studying choice problems in which the attributes are inter-correlated. In contrast, regression analysis affords some solution to this problem, although multi-collinearity difficulties can arise. Within the ANOVA approach, elements of  $\underline{S}$  must be described in a categorical fashion; whereas regression analysis allows members of  $\underline{S}$  to be specified categorically, along a continuum, or as a combination of the two. The multiple regression methodology is normally carried out on the individual subject level, in contrast to the group analyses usually performed in the ANOVA approach. Also, the statistical significance of the estimated weight parameters is rarely tested in regression analyses, whereas the statistical significance of the degree of importance of each dimension is tested as a matter of course in ANOVA studies.

It should be noted that the last two differences mentioned above reflect only a divergence in interests of those applying the techniques and are not in any way an indication of limitations in either



the regression or ANOVA paradigms. Although ANOVA and multiple regression are formally equivalent models for assessing weights of importance dimensions, it is probably useful to maintain the distinction between the two, if only for historical reasons.

The procedures described in the following sections are all special cases of the ANOVA methodology, each requiring substantially fewer holistic evaluations than needed to perform the complete factorial ANOVA just discussed. The last two approaches, Holistic Orthogonal Parameter Estimation (HOPE) and the indifference procedures, require an extremely small number of judgments, as is characteristic of the direct subjective techniques previously described. Within the present taxonomy of approaches to importance weighting, however, these two techniques must be categorized as indirect holistic procedures: the subject is required to make judgments directly relating to elements of  $\underline{S}$  and not to any abstraction, such as a dimension or attribute.

#### Fractional Replication Designs for ANOVA

For a complete factorial ANOVA design, the subject must provide holistic evaluation of  $\underline{a}_1 \times \underline{a}_2 \times \dots \times \underline{a}_k \times \underline{r}$  different elements of  $\underline{S}$  (where  $\underline{a}_i$  is the number of levels of the factor  $\underline{A}_i$ ,  $\underline{k}$  is the total number of factors, and  $\underline{r}$  is the number of times each stimulus is judged). When  $\underline{k}$  is large or many of the  $\underline{a}_i$  are large, the number of holistic evaluations required will become unwieldy. One solution to this problem is simply to replace the complete factorial design with a fractional replication design. (For details of how to construct such designs, see Cochran and Cox, 1957). Slovic (1969) used such a

procedure in a study designed to determine the factors upon which stockbrokers rely when evaluating prospective stock purchases. Faced with the prospect of getting evaluations of 2048 companies (11 factors, 2 levels of each,  $2^{11} = 2048$ ) from professional stock brokers, Slovic elected to use a 1/16 fractional replication, requiring evaluations of only 128 companies. In another study involving judgments from stock brokers and MBA students, Slovic, Fleissner, and Bauman (1972) reduced a problem requiring 256 evaluations (8 factors, 2 levels of each,  $2^8 = 256$ ) to one involving only 64 by using a 1/4 fractional replication.

The only assumption that must be made to justify the use of fractional replication designs is that interaction effects involving three or more factors are negligible. If this assumption is justified, then the fractional replication strategy affords an extremely "cheap" method of assessing the strength of main effects and two-way interactions. Since the higher-order interaction terms are used as estimates of the error term, the penalty for utilizing the procedure when higher-order interactions are large is to inflate the error term, thus underestimating the size of the main effects (weights) and two-way interactions. Procedures for obtaining weight estimates are identical to those described for the ANOVA paradigm.

#### Holistic Orthogonal Parameter Estimation

Barron and Person (in press) proposed a general methodology for eliciting multi-attribute models via holistic evaluations of choice alternatives. Within the HOPE procedures, one can obtain estimates of the weight parameters in Equation 2 for riskless, additive multi-attribute utility functions. Similar to the strategy employed in the

fractional replication approach, the HOPE procedures requires evaluations of only a subset of the elements needed for a complete factorial design. In ANOVA terminology, HOPE requires evaluations of subsets of  $\underline{S}$  necessary to create an orthogonal design. (For details of how to construct such sets of alternatives, see Addelman, 1962). In one example, Barron and Person (in press) constructed fifteen elements of  $\underline{S}$  needed for an orthogonal design, to be used in place of a complete factorial design with five factors, each defined on four levels ( $4^5 = 1024$  for the complete factorial).

The weights in Equation 2 are estimated in a manner completely analogous to standard ANOVA weight estimation procedures.  $\underline{W}_i$  is calculated as the difference between the mean evaluation of all alternatives which are best on dimension  $\underline{i}$  and the mean evaluation of all alternatives worst on dimension  $\underline{i}$ . (Of course, this requires that the nominal levels "best" and "worst" for each dimension can be independently identified.) This computational procedure is precisely equivalent to that demonstrated in Kerlinger and Pedhazur (1973) for determining regression weights of "dummy coded" predictor variables. Thus, Barron's HOPE procedures can be viewed as an ingenious adaptation of both the bootstrapping and functional measurement paradigms.

#### Indifference Techniques

Two techniques, willingness to pay (pricing out) and trading off to the most important dimension (trade off), have been developed for the assessment of weight parameters (Keeney and Raiffa, 1976; MacCrimmon, 1973). In the pricing-out method, the subject must state an amount

(usually in dollars) which represents the difference in value between two choice entities, identical on all dimensions except one. On that single discrepant dimension, one of the alternatives is defined as best and the other as worst. By choosing the alternative pairs such that the discrepant dimension varies, the  $k$  numbers (prices) are obtained. The weight on the  $i$ th dimension is simply the price (normalized by dividing by the sum of all the elicited prices) stated as the difference between alternatives worst and best on attribute  $i$ , given equivalence on all other attributes.

In the trade-off method, one begins by determining which of the  $k$  alternatives, each defined as best on one dimension and worst on all others, is preferred. Let  $p$  be the dimension upon which the most preferred alternative is best. One then determines alternatives which are worst on all dimensions except dimension  $p$  and equivalent in value to each of the  $k-1$  non-preferred alternatives mentioned before. Thus, the subject specifies the amount of change on the most important dimension equivalent to a change from worst to best on each of the other dimensions. The weight on the  $i$ th dimension is defined as the value on dimension  $p$  needed to make an alternative worst on all dimensions, except  $p$ , equivalent to an alternative best on dimension  $i$  and worst on all other dimensions.

The relationship between the indifference methods and the more statistically-oriented approaches of ANOVA, dummy-coded regression analysis, and HOPE should be apparent. Rather than obtaining estimates of the differences between best and worst levels on each dimension by collecting several holistic evaluations, the pricing-out and trade-off procedures require that subjects directly estimate these differences,



in terms of either dollars (pricing-out) or value on the most important dimension (trade-off). (One can always think of pricing out as a trade-off procedure in which money is the most important dimension.) The weights are then derived from these difference estimates in the same fashion as in the more statistical procedures.

One criticism of the pricing methodology is the potential difficulty in specifying differences between worst and best on all attributes in terms of money. For some decision problems, changes from best to worst on attributes are not readily conceptualized in terms of financial loss. Of course, the same criticism can be made of the trade-off procedure, in the sense that it may not be convenient to think of changes on some attributes in terms of changes on the dimension deemed most important. An additional problem with the trade-off procedure concerns the difficulties involved in trying to conceive of choice alternatives which are "best on one dimension and worst on all others." The complete implausibility that such an element of  $S$  could exist may make the procedure totally useless for some weight assessment problems.

### Convergent Validation of Importance Weights

A fair amount of empirical research has been devoted to comparing various methods of obtaining importance weights, yet there is presently no strong evidence about which assessment procedure produces a more accurate estimate of weight parameters in additive, riskless multi-attribute utility functions. Every one of the weighting techniques reviewed above is currently in vogue, both in "descriptive" laboratory research on human judgment and in "prescriptive" applications of the normative technology of decision-making. Of course, this diversity may be attributed, in part, to the fact that the efficacy of a given assessment procedure is a function of the specific decision problem. Even a casual perusal of the literature on eliciting importance weights, however, reveals that very little of the variance in the assessment procedure employed is accounted for by the defining characteristics of the situation. Instead, one finds that each researcher and/or decision analyst consistently employs a single methodology, largely independent of the specific characteristics and demands of the decision problem. As long as the various procedures are all equally accurate at weight parameter estimation, there is nothing wrong with this state of affairs. But these methods vary in ease of use and in amount and type of information obtained. So they are unlikely to be all equally accurate, and if they were, they are certainly not equally attractive.

The research devised to compare various weighting procedures falls into three major categories, corresponding to three different definitions of importance weight validity: (1) correspondence of actual weight estimates, (2) correspondence of composites derived from

estimated weights and holistic evaluations of alternatives, and (3) correspondence of composites derived from different weight estimates. A review of the convergent validity research from each of these three perspectives follows.

#### Correspondence of Actual Weight Estimates

Perhaps the most straightforward tactic for comparing different weighting methodologies is simply to examine the assessed weights themselves. Unfortunately, little attention has been devoted to developing a good direct measure of correspondence between sets of weights. Most studies, in fact, have only reported sets of assessed weights, either in numerical or graphical form. In some instances a correlation coefficient has been computed across pairs of weighting schemes, but the problem of such a criterion of correspondence between importance weights has been discussed earlier. The correlation is only unique up to a linear transformation, which is simply not a sensitive direct measure of correspondence between weighting schemes.

Hoffman (1960) reported one of the first studies comparing a direct subjective procedures with an indirect holistic one. He found relatively good correspondence between "relative weights" ( $\hat{\beta} \cdot \underline{r}$ , obtained from a regression analysis of holistic evaluation) and weights determined from an indirect subjective assessment, fractionation, for two different decision problems (evaluating intelligence, defined on nine dimensions, and sociability, defined on eight dimensions). The distributions of subjective weights were somewhat more uniform than those for relative weights. Since Hoffman only presented the graphic plots of weight pairs for selected subjects, no conclusion beyond a simple subjective impression that the weight profiles seem essentially

to agree is possible.

In a study of two stock brokers' preference functions for various stocks defined on 11 dichotomous dimension, Slovic (1969) reported a comparison among three different weighting schemes, two derived statistically from a fractional replication ANOVA design and one from the subjective 100-point fractionation method. The weight profiles of the "magnitude of effect" index and the directly assessed weights were in close agreement; however, the index of proportion of variance accounted for,  $\omega^2$ , yielded weights which were more extreme than those from the other two methods. Since  $\omega^2$  is a simple function of the squares of the magnitude-of-effect statistic, this result is not surprising. Again, only the graphic plots of weight profiles were presented; no measure of their agreement is given.

Eckenrode (1965) compared six different methodologies for obtaining subjective importance weights and found little difference among them. Three separate problems, using either five or six dimensions, yielded correlations among mean weight assessments in the high nineties (most over .975). Unfortunately, this nomothetic analysis conveys little information concerning the relationship of individual subjects' weights obtained via different elicitation procedures.

In a study of future socio-economic growth of underdeveloped nations (current status specified on four attributes), Summers et al. (1970) found that the mean relative weights ( $\hat{\beta} \cdot \underline{r}$ ) were less uniform than subjective weights obtained from a direct 100-point fractionation procedure. An individual subject analysis also revealed that significant beta weights were obtained on far fewer dimensions than subjects directly reported using.



Using a somewhat unorthodox methodology, Blood (1971) obtained measures of satisfaction with five different dimensions of job situation, an overall measure of job satisfaction, and a rank-ordering of the five job dimensions from 380 clerical workers. One set of numbers was obtained for each subject, with the job situation of relevance being their own. After reordering each subject's dimensions from most important to least, Blood computed statistical measures of importance (betas, validity coefficients, and the Usefulness measure) for the five dimensions, now defined not by content area, but by the idiosyncratically assigned importance rank of each subject. The results of the regression analysis indicated no relationship between the subject's rankings of the attributes of job satisfaction and any set of statistical importance weights. Although Blood interpreted these results as evidence against the validity of direct weight assessments, there is an alternative explanation. Since Blood obtained data for each different job situation from a different subject, the measures of job satisfaction, both for the five attributes and the overall, may not be comparable across subjects. Of course, it only makes sense to compute statistical measures of importance if the meaning of the scale values is invariant across all 380 subjects, each of whom contributed one row of the data matrix. For example, one of the more objective scales was "pay." It is highly unlikely that all 380 subjects have the same utility function for money: an unmarried mother of four is likely to view \$1000 per month salary quite differently from a woman married to a physician who makes \$80,000 per year. Without some evidence that the numbers obtained mean the same thing to all subjects, Blood's conclusions must be treated with caution.

One of the more curious findings was reported by Slovic et al. (1972) in a study of stock brokers' and students' utility functions for stock investments. Slovic et al. found that the subjective weights of stock brokers over the eight dichotomous dimensions (derived via the 100-point fractionation method) did not correspond closely to the magnitude-of-effect measures derived from the ANOVA. The correlation of stock brokers' mean weights for the two schemes was .34, while the subjective and statistical weights of the students were in much greater agreement, yielding a correlation between the mean weights across students of .79. Slovic hypothesized that the recent learning experiences of the students led to a greater awareness of the mechanics of the evaluation process, whereas the more experienced stock brokers were used to making more or less automatic evaluations requiring little attention.

Permut (1973) compared a subjective ranking of the importance of ten dimensions of instructor quality with the beta weights derived from a regression analysis based on holistic evaluations of hypothetical instructors. Across fourteen subjects, Permut reported a mean rank-order correlation between the two weighting schemes of .55 (values ranged from -.13 to .84, with eight of the fourteen significantly greater than zero). One suspicious result of the Permut study is the prevalence of negative statistical weights for dimensions which are presumed to be positively correlated with overall instructor quality. This suggests the possibility of multi-collinearity problems among the attributes, and the potential for poorly estimated regression weights cannot be overlooked. Since Permut does not report the intercorrelations

among the ten attributes, this hypothesis is only a speculation about why some of the subjective and regression weights are discrepant.

Aside from the technical problems already mentioned in using a direct comparison of weight estimates, there is a much more compelling reason to seek another strategy for validating weight estimates. In general, one is not interested in the correspondence of decision rules. Instead, the topic of interest is really the correspondence of the evaluations produced by those decision rules (Edwards, Note 7). With this principle in mind, it seems reasonable to re-focus attention away from a comparison of weight assessments and toward the composites derived from the weight assessments.

#### Correspondence of Composites Derived from Assessed Weights and Holistic Evaluations

A number of studies have compared various weighting schemes by simply examining the extent to which each is able to predict holistic evaluations of a set of multi-attributed stimuli. Unfortunately, virtually all such comparisons involve holistic evaluations used to generate the statistical weights. Although the multiple correlations involving composites formed from regression weights are sometimes adjusted downward to account for the well-known "over-prediction" problem, a better approach would be to cross-validate the statistical weighting scheme with holistic evaluation not used in estimating the regression weights. Few studies have bothered to take such precautions.

Summers et al. (1970) report a drop in median multiple correlation from .75, for optimal beta weights, to .60 for direct estimates of the importance weights. Since Summers et al. do not report adjusted multiple correlations nor cross-validated multiple correlations, this drop in predictive power is difficult to interpret.

Hoepfl and Huber (1970) found little difference in the predictive power of regression weights and directly estimated weights for attributes of instructor quality. For six problems involving from one to six attributes, they found that the median multiple correlation, across subjects, ranged from .87 to .98 for the direct subjective weight estimates, whereas median multiple correlations (adjusted for inflation) ranged from .91 to .98. The median multiple correlation for the subjective weights was actually larger than the median adjusted multiple correlation obtained from regression analysis for two of the six problems.

A rather important finding is reported by Huber, Daneshgar, and Ford (1971) in a field study of job preference. Using job descriptions specified on five dimensions, Huber et al. found that the predictive power of regression weights and subjective weights varied substantially as a function of the actual job experience of the subject. Although experienced subjects' holistic evaluations were more accurately predicted by a model based on optimal regression weights (mean adjusted  $\bar{R} = .80$  versus the subjective weight model mean  $\bar{R} = .62$ ), inexperienced subjects' evaluations were much closer to the composite resulting from the subjectively estimated weights (subjective weight model mean  $\bar{R} = .67$  versus mean adjusted  $\bar{R} = .41$ ). This finding, somewhat similar in spirit to that of Slovic et al. (1972), supports the notion that less experienced subjects may be better able to report subjective estimates of model weight parameters.

Nystedt and Magnusson (1975) also compared composites resulting from regression weights to those derived from subjective weights for six subjects and three different clinical evaluation tasks. Being quite careful to use a double cross-validation procedure for estimating



the multiple correlations resulting from regression analysis, they found that holistic evaluations were much closer to subjective weight composites than to regression weight composites. For all three tasks, the average multiple correlation for the subjective weight model was higher than that obtained upon a cross-validation of the optimal regression weight model. Unfortunately, the results of the Nystedt and Magnusson study are difficult to interpret because of the somewhat unusual manner in which the subjective weight model was constructed. Specifically, Nystedt and Magnusson chose to estimate different sets of subjective weights for each of the choice alternatives. That is, the weight parameters in Equation 2 were re-estimated for every element of  $S$ , and the overall composite for each element was determined from weights estimated for that element alone. Little in this model and assessment procedure seems relevant.

In a study comparing seven different subjective weighting schemes and weights obtained from a regression analysis of holistic choices of financial aid applicants (defined on three dimensions) and graduate admissions applicants (defined on seven dimensions), Cook and Stewart (1975) found little difference in the predictive validity of the eight sets of weights. The ratio of the mean (across subjects) squared multiple correlation of the subjective weight models to the unadjusted squared multiple correlation of the optimal regression model ranged from .88 to .95, for the three-attribute problem, and from .68 to .79, for the seven-attribute problem. For both problems, the 100-point fractionation procedure yielded the highest average ratio. The average squared correlation over the seven subjective weighting methodologies was .74 for the financial aid task and .53

for the graduate admissions problem. In comparison, the squared multiple correlation corrected for shrinkage was .80 for the financial aid problem, and the baseline correlation resulting from equal weighting was .62. For the graduate admission problem, these respective correlations were .66 and .39, for optimal beta and equal weights. Thus, although the subjective weights represented a fairly substantial increase in predictive validity over equal weights, they did not quite match that of the optimal beta weights. Again, the authors offered no cross-validation of the weights derived via regression analysis.

von Winterfeldt and Edwards (Note 2) found that correlations between decomposed composite evaluations of apartments and holistic evaluations ranged from .488 to .819, across four subjects. The somewhat lower correlations reported in this study may be explained by the fact that the apartments were defined on fourteen dimensions of importance, far more than studied in any of the other investigations discussed above.

The fundamental issue in studying the validity of parameter estimates by comparing the derived composites with holistic evaluations is an obvious one. If holistic evaluations are a valid criterion, what need is there for a model of any kind? An ordering of  $S$  could be obtained from simple holistic evaluations of the alternatives, and no complicated assessment procedures would be required. The problem, of course, lies in the fallibility of holistic evaluations. A large body of research suggests that holistic choices and appraisals are easy to "out-predict" (for a review, see Slovic & Lichtenstein, 1971). Human judges, even trained experts, do not do well when aggregating several pieces of information simultaneously. The rationale for statistical

weighting schemes derived via holistic choices is to eliminate the random noise and inconsistency present in raw holistic evaluations. The entire logic of the direct subjective decomposition approach, on the other hand, depends upon the edge afforded by the "divide and conquer" strategy, allowing the decision-maker to avoid the problems of aggregating information and to concentrate on less complicated questions of preference. Thus, there seems to be some need for a measure of validity other than that provided by the prediction of holistic choice.

#### Correspondence of Composites Resulting from Different Weighting Procedures

One approach, not often taken, has been to calculate the correlations of composites derived from different weight assessment procedures. If these correlations are high, then one can argue for evidence of convergent validity. That is, a high correspondence of composites across assessment procedures suggests that the construct of multi-attribute utility is real and measurable. Unfortunately, researchers have rarely, if ever, tried to test for the converse of convergent validity, i.e., discriminant validity, in the sense of demonstrating low correlations between evaluations of different alternative sets obtained from the same assessment procedure. The logic is roughly the same as that proposed in the "multi-trait, multi-method" approach of Campbell and Fiske (1959).

An example of convergent validation is Fischer and Peterson (Note 8). They correlated the composites formed from weights derived by magnitude estimation versus ratio estimation of the importance of six dimensions of instructors' teaching ability. Over sixteen subjects,

correlations between the two approaches ranged from .78 to 1.0, with a median correlation of .92.

A number of studies have compared composites derived from direct estimates of the weight parameters in Equation 2, with composites formed from risky assessment procedures. Since models and assessment methodologies other than those of riskless, additive utility are not the focus of the current discussion, the details of the already mentioned studies by Fischer (1976, 1977, Note 1) and von Winterfeldt and Edwards (Note 2) will be omitted. Although some discrepancies between the risky and riskless procedures are evident, the overwhelming result is that of convergence.

A rather interesting approach was taken by Klahr (1969). The logic of his idea is elegantly simple: (1) obtain similarity judgments of pairs of the  $\underline{S}$  elements; (2) perform a multi-dimensional scaling on these similarities, resulting in the specification of each alternative as a vector in an Euclidean space of some dimensionality (The number of dimensions required is usually determined by a criterion, set by the experimenter to describe the goodness of fit of the similarity data to the resulting spatial representation.); (3) locate an "ideal" point in the space using some independent procedure (This point may be an actual member of  $\underline{S}$ , or only hypothetical.); (4) calculate distances, within the  $\underline{S}$ -space, between every point in  $\underline{S}$  and the "ideal" point; (5) for any set of composites resulting from a decomposition assessment procedure, compare the distances computed in step 4 with the overall value of the composites. Naturally one would be delighted to find that such a comparison yielded a perfectly inverse relationship, since, intuitively, overall utility and distance from an



ideal point are highly related to the same construct. Indeed, Klahr (1969) performed exactly the steps outlined above and obtained excellent convergence.

Clearly, all of the previously discussed work on equal weights and rank weights falls into the category of correlating composites from different weighting schemes. Virtually all of this research, however, suffers from the same flaw, originally detected by Seaver (Note 4) and formalized by McClelland (in press): correlations are frequently calculated over sets including alternatives not part of the Pareto optimal frontier. McClelland has shown the seriousness of this problem for Pareto frontiers in choice situations characterized as "pick one out of  $n$ ." However, there has been no discussion of the topic, first recognized by Seaver (Note 4), of the more general problem characterized as "pick  $k$  out of  $n$ ." As Seaver points out in the context of MAU validation: "if more than one alternative is to be selected, use only of the admissible set is not appropriate (Note 4, p. 6)."

### Criterion Validation of Estimated Weight Parameters

Ultimately, evidence of convergent validity can only provide partial satisfaction for anyone seriously interested in justifying estimates of weight parameters. One would like a criterion against which to compare weight estimates and the composites derived from them. In the past, utility has usually been treated as a subjective construct, which, like other internal constructs such as attitude, cannot be right or wrong. Unlike subjective probabilities, subjective values have traditionally been considered beyond the scope of criterion validation, because no convenient criterion, such as relative frequency, existed.

The field study by Huber et al. (1971) utilized the job eventually chosen by the applicant as an external validity criterion against which to compare different model predictions. The criterion measure considered was the ranking by each model of the actual job which applicants eventually took. Although one might argue that the job actually obtained is simply another form of holistic judgment, we feel that this is only partly true. The subjects all lived with their job choices after the experiment was over. The job chosen is a strictly behavioral measure, in contrast to the simple cognitive measure of holistic evaluation. Thus, the job obtained seems closer to an objective criterion than to a prototypical holistic evaluation.

The subjective weights were far more predictive than the statistical ones, identifying the job actually chosen as the most preferred for 10 of the 15 experienced subjects and 8 of the 15 inexperienced, as compared to 7 of the 15 experienced and 2 of the 15 inexperienced subjects for the

regression analysis model. In addition, the ranks of the chosen jobs which were not identified as most preferred were generally much higher for the subjective-weight model than for the beta-weight model. The success of subjective weights in predicting actual preferences in a nonlaboratory situation is strong evidence for their usefulness in estimating the weight parameters defined in Equation 2.

Huber et al. (1971) point out several aspects of the field conditions which could have added a considerable amount of error variance to the data. For one thing, only jobs which were "acceptable" to each applicant were considered in the analysis. It is possible that a job that was holistically considered unacceptable would have received high model scores, thus changing the ranking of the job that was actually chosen. Second, most of the salaries for the jobs, one of the five attributes included in the model, were not known precisely; estimates were based on the "book" salary of the job. Third, and perhaps most important, is the fact that the job desired and the job obtained were often not the same. Many of the jobs were filled before candidates could schedule an interview, and many of the candidates probably made poor impressions at interviews for jobs which were desired and hence were not hired. Fourth, some of the job choices were probably determined by factors not included in the five-factor model used to make the predictions. Finally, the interview experiences between the time the model parameters were estimated and the time jobs were actually accepted may have significantly modified the candidate's preference model. In all, it is quite surprising that the models performed as well as they did. However, one might entertain the possibility that the superiority of directly-assessed subjective weights over the

statistical weights is, in part, due to an interaction of model robustness with the various extraneous variables mentioned above.

Within this same framework, Oskamp (1967) reported a study of clinical diagnosis using the MMPI. Oskamp had 21 Ph.D-level clinicians and 24 graduate students rate 200 MMPI profiles for whether the patient was being hospitalized for psychiatric or medical reasons. (Half of these were re-evaluated after one month to assess reliability.) Statistical weights on each of the 13 MMPI scales were obtained from each judge's holistic evaluations (Normalized "relative weights" --  $\hat{\beta} \cdot \underline{r} / R^2$  -- were computed, as well as normalized validity coefficients ---  $\underline{r} / \Sigma \underline{r}$ .) Subjective weights were directly assessed for each judge as a "subjective report of what proportional weight he thought he was attaching to each variable (Oskamp, 1967, p.412)." Since the profiles were of actual patients, "true" statistical weights were obtained by using the dichotomy of actual diagnosis (psychiatric vs. medical) as a criterion variable. Again, both relative weights and normalized validity coefficients were calculated.

Unfortunately, the critical analyses comparing the composites resulting from the various weighting schemes were not reported. However, the average number of scales which each judge reported using (subjective weights) matched the average number of significant scale validity coefficients determined from the indirect holistic weight assessment. Both of these numbers were close to the number of significant correlations between the MMPI scales and the criterion diagnosis. Furthermore, the ordering and relative magnitudes of both the directly-assessed subjective weights and the validity coefficients derived from holistic assessment matched the "true" weights (validity coefficients) closely and to approximately the



same degree. In general, the subjective weights derived were quite close to both sets of validity coefficients:

Weight analysis suggests that (a) the Js knew what they were doing and reported on their decision processes quite accurately, and (b) they used just about the optimum amount of complexity in their decisions (Oskamp, 1967, p. 413).

The analyses involving relative weights were quite inconsistent with those using validity coefficients. Part of the problem may lie in the high multi-collinearity among the 13 MMPI scales. Although no scale intercorrelations are reported, Oskamp (1967) stated that many of the pairs of scales are "highly intercorrelated, such as the MMPI Pt and Sc scales (p. 414)." The MMPI Handbook, Vol. 2 (Dahlstrom, Welsh, & Dahlstrom, 1975) gives typical intercorrelations between Psychasthenia (Pt) and Schizophrenia (Sc) in the mid to high 80's. Thus, although sufficient data are not presented to make a definite statement, it is quite likely that the regression weights were poorly estimated using the standard least-squares approach, and that the analyses using the derived "relative weights" should be discounted.

Brehmer and Qvarnstrom (1976) used an ingenious paradigm for testing the validity of subjective weighting techniques." They provided subjects with "true" ratios of importance weights (two attributes only) in a MCPL task. By calculating the statistical indices of importance derived from the holistic evaluations, Brehmer and Qvarnstrom (1976) found:

that the subjects have an intuitive understanding of the concept of weight, and that their understanding of the weight concept corresponds to the slopes of

the functions relating the judgments to the cue values, rather than to the variance accounted for by the cues (p. 125).

If stimulus weights are interpreted by subjects as correlations (or slopes) and not as "proportion of variance accounted for," then subjective weights given as responses by subjects should probably be interpreted as correlations or slopes, also. As Brehmer and Qvarnstrom point out, the lack of correspondence between subjective and statistical weights found by Summers, et al. (1970) and by Slovic (1969) can be attributed to the specific type of statistical measure of importance used.

If, however, the objective weights are redefined in terms of slope coefficients, a close correspondence between objective and subjective weights is obtained, indicating that the subjects did, in fact, know their own policies (Brehmer & Qvarnstrom, 1976, p. 125).

In another study, also in the MCPL framework, Schmitt, Coyle, and Saari (1977) investigated how subjects interpret so-called "task information" feedback. In two studies, each involving three cues, they provided either (1) no task information, (2) "true" validity coefficients only ( $r$ ), (3) "true" regression weights only, or (4) "true" relative weights only ( $\hat{\beta} \cdot r$ ). Thus, the experiment is exactly analogous to that performed by Brehmer and Qvarnstrom (1976), except that the subjects are given the "true weights" in terms of a variety of different statistical measures of importance. Schmitt et al. (1977) found that correlations between composites derived from subject's holistic evaluations and

composites formed from the true model were higher for subjects given either  $r$  or  $\hat{\beta}$  task information than for subjects given  $\hat{\beta} \cdot r$  feedback. Thus, further evidence is provided for the hypothesis that subjects interpret the concept of importance in terms of correlation or slope, and not in terms of "variance accounted for."

An important recent study by Schmitt (1978) provides strong evidence that subjects can, in fact, give direct assessments of attribute importance. In a MCPL setting, subjects evaluated students' expected academic performance (GPA) on the basis of either three or four of the following attributes (test scores): quantitative ability, verbal ability, responsibility, and past academic achievement. Composites were formed based on four sets of estimated importance weights: (1) regression weights derived from holistic judgments, (2) 100-point fractionation, (3) ten-point rating scale, and (4) ratio estimation. The important result from Schmitt's experiment involves the "matching" index, or the degree to which composites formed from the various weighting procedures correlate with composites formed from the "optimal" (regression) weights derived from the outcome feedback criterion. As Schmitt (1978) argues, "matching represents the degree to which the subject 'knows' the real relationship among the variables (p. 176)." Across all experimental conditions (three attributes vs. four, first trial block vs. second, and outcome feedback vs. no outcome feedback), Schmitt found that there were absolutely no differences between the matching indices produced from the four weighting procedures. The convergence of all four importance weight estimates gives further support to the high convergence found by Cook and Stewart (1975) over seven different weighting procedures. However, because outcome feedback in the MCPL paradigm defines a set of

"true" weights, stronger conclusions are warranted. The high degree of matching reported in this study suggests that subjects' weight estimates were not only convergent, but accurate; i.e., criterion validity was demonstrated.

Two additional points about Schmitt's study should be made. First, the intercorrelation matrix for the four attributes contained two rather large correlations, both .78. For each of the two subsets of three attributes used, one of the large (.78) correlations was also present. The high level of multi-collinearity may have led to serious mis-estimation of the "optimal" weights. Although all of the "true" validity coefficients ranged from .42 to .53, the least-squares estimates of the regression weights were non-uniform and negative beta weights resulted (e.g., .63, -.15, .16, .40, for the four-attribute problem). The second point is related. The average subjective weights, reported by Schmitt for each experimental condition, are markedly uniform. (The maximum ratio for any two weights in each set is about 2, and most are essentially equal.) Thus, his subjects' subjective weights more closely correspond to the validity coefficients of the attributes than to the least-squares regression weights, which, as was pointed out above, were quite non-uniform.



### Conclusions and Discussion

In summary, the weighting literature reviewed, and particularly the recent criterion validation work, suggests that the concept of attribute importance is a psychologically meaningful one. For many of the laboratory and field settings studied, subjects gave responses to direct subjective assessments of importance weights that were both consistent (high convergent validity) and accurate (high criterion validity). Few discrepancies were observed in studies comparing direct subjective estimates of importance to statistical indices of importance derived indirectly from holistic evaluations. Furthermore, there is some evidence to indicate that the psychological concept of importance is more closely related to the statistical notion of attribute validity (correlation) than to either the least-squares regression weights (Schmitt, 1978) or the "proportion of variance accounted for" (Brehmer and Qvarnstrom, 1976; Schmitt et al., 1977; Slovic, 1969).

The notion that judges are aware of the importance that they attach to attributes, or that should be attached to attributes, is contradictory to the conclusions reached by Slovic and Lichtenstein (1971). In their review of regression approaches to judgment, Slovic and Lichtenstein (1971) "found serious discrepancies between subjective and computed relative weights (p. 684)". This often quoted conclusion (see Nisbett & Wilson, 1977) has become known as the "self insight error," and is probably responsible for the relative paucity of research on subjective weights as compared to that on statistical weights derived from holistic evaluations.

## REFERENCE NOTES

1. Fischer, G. Four methods for assessing multi-attribute utilities: An experimental validation (Tech. Rep. 037270-6-T). Ann Arbor, Mich.: University of Michigan, Engineering Psychology Laboratory, June, 1972.
2. von Winterfeldt, D., and Edwards, W. Evaluation of complex stimuli using multi-attribute utility procedures (Tech. Rep. 011313-2-T). Ann Arbor, Mich.: University of Michigan, Engineering Psychology Laboratory, October, 1973.
3. Newman, J. R., Seaver, D. A., and Edwards, W. Unit versus differential weighting schemes for decision making: A method of study and some preliminary results (SSRI Tech. Rep. 76-5). Los Angeles: University of Southern California, Social Science Research Institute, July, 1976.
4. Seaver, D. A. Correlational analyses of approximations and sensitivity for multi-attribute utility. Unpublished manuscript, 1973. (Available from Decision Science Consortium, Inc., Reston, Virginia.)
5. Stillwell, W., and Edwards, W. Equal weights versus rank weights: Prediction versus decision. Unpublished manuscript, 1978. (Available from Social Science Research Institute, University of Southern California, Los Angeles.)
6. Eils, L. C., III, and John, R. S. Multi-attribute utility analysis and communication strategy in group decision-making: A criterion-related validity experiment (SSRI Tech. Rep. 78-4). Los Angeles: University of Southern California, Social Science Research Institute, December, 1978.
7. Edwards, W. Multiattribute utility measurement in a highly political context: Evaluating desegregation plans in Los Angeles. Unpublished manuscript, 1978. (Available from Social Science Research Institute, University of Southern California, Los Angeles.)

8. Fischer, G. W., and Peterson, C. R. Ratio versus magnitude estimates of importance factors (Tech. Rep. 037230-3-T). Ann Arbor, Mich.: University of Michigan, Engineering Psychology Laboratory, June, 1972.
9. Otway, H., and Edwards, W. Application of a simple multiattribute rating technique to evaluation of nuclear waste disposal sites: A demonstration (Research Memorandum RM-77-31). Laxenburg, Austria: International Institute for Applied Systems Analysis, June, 1977.

## REFERENCES

- Abelson, R. P., and Tukey, J. W. Efficient utilization of non-numerical information in quantitative analysis: General theory and the case of simple order. Annals of Mathematical Statistics, 1963, 34, 1347-1369.
- Addelman, S. Orthogonal main effect plans for asymmetrical factorial experiments. Technometrics, 1962, 4, 2-46.
- Anderson, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes (Eds.), Contemporary developments in mathematical psychology (Vol. 2). San Francisco: Freeman, 1974.
- Anderson, N. H. Note on functional measurement and data analysis. Perception and Psychophysics, 1977, 21, 201-215.
- Barron, H. F., and Person, H. B. Assessment of multiplicative utility functions via holistic judgments. Organizational Behavior and Human Performance, 1979, 20 (in press).
- Blood, M. R. The validity of importance. Journal of Applied Psychology, 1971, 55, 487-488.
- Brehmer, B., and Qvarnstrom, G. Information integration and subjective weights in multiple-cue judgments. Organizational Behavior and Human Performance, 1976, 17, 118-126.
- Campbell, P. T., and Fiske, D. W. Convergent and discriminant validation by the multitrait multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.
- Cochran, W. G., and Cox, G. M. Experimental designs (2nd ed.). New York: Wiley, 1957.
- Cohen, J. Multiple regression as a general data-analytic system. Psychological Bulletin, 1968, 70, 426-443.



- Comrey, A. L. A proposed method for absolute ratio scaling. Psychometrika, 1950, 15, 317-325.
- Cook, R. L., and Stewart, R. R. A comparison of seven methods for obtaining subjective descriptions of judgmental policy. Organizational Behavior and Human Performance, 1975, 13, 31-45.
- Dahlstrom, W. G., Welsh, G. S., and Dahlstrom, L. E. An MMPI handbook. Volume II: Research applications. (Rev. ed.) Minneapolis: University of Minnesota Press, 1975.
- Darlington, R. B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69, 161-182.
- Darlington, R. B. Reduced-variance regression. Psychological Bulletin, 1978, 85, 1238-1255.
- Dawes, R. M. A case study of graduate admissions: Application of three principles of human decision making. American Psychologist, 1971, 26, 180-188.
- Dawes, R. M., and Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 81, 95-106.
- Dempster, A. P., Schatzoff, M., and Wermuth, N. A simulation study of alternatives to ordinary least squares. Journal of the American Statistical Association, 1977, 72, 77-91.
- Eckenrode, R. T. Weighting multiple criteria. Management Science, 1965, 12, 180-192.
- Edwards, W. Social utilities. The Engineering Economist, Summer Symposium Series, 1972, 119-129.
- Edwards, W. How to use multiattribute utility measurement for social decision making. IEEE Transactions on Systems, Man, and Cybernetics, 1977, SMC 7, 326-340.

- Einhorn, H. J., and Hogarth, R. M. Unit weighting schemes for decision making. Organizational Behavior and Human Performance, 1975, 13, 171-192.
- Fischer, G. W. Multidimensional utility models for risky and riskless decisions. Organizational Behavior and Human Performance, 1976, 17, 127-146.
- Fischer, G. W. Convergent validation of decomposed multiattribute utility assessment procedures for risky and riskless choice. Organizational Behavior and Human Performance, 1977, 18, 295-315.
- Fujii-Eustace, T. The reliability and validity of multi-attribute utility measurement methods: An experimental comparison. Unpublished doctoral dissertation, University of Southern California, 1978.
- Gardiner, P. C., and Edwards, W. Public values: Multiattribute utility measurement for social decision making. In M. F. Kaplan and S. Schwartz (Eds.), Human judgment and decision processes. New York: Academic Press, 1975.
- Ghiselli, E. E. Theory of psychological measurement. New York: McGraw-Hill, 1964.
- Goldberg, L. R. Diagnosticians vs. diagnostic signs: The diagnosis of psychosis vs. neurosis from the MMPI. Psychological Monographs, 1965, 79, (9, Whole No. 602).
- Goldberg, L. R. Simple models or simple processes? Some research on clinical judgments. American Psychologist, 1968, 23, 483-496.
- Goldberg, L. R. Man versus model of man: A rationale plus some evidence for a method of improving on clinical inferences. Psychological Bulletin, 1970, 73, 422-432.
- Goldberg, L. R. Five models of clinical judgment: An empirical comparison between linear and nonlinear representations of the human inference process. Organizational Behavior and Human Performance, 1971, 6, 458-479.

- Gulliksen, H. Theory of mental tests. New York: John Wiley & Sons, 1950.
- Guttentag, M., and Snapper, K. Plans, evaluations, and decisions. Evaluation, 1974, 2, 58-74.
- Hammond, K. R. Probabilistic functionalism: Egon Brunswik's integration of the history, theory, and method of psychology. In K. R. Hammond (Ed.), The psychology of Egon Brunswik. New York: Holt, Rinehart, & Winston, 1966.
- Hammond, K. R., Stewart, T. R., Brehmer, B., and Steinmann, D. O. Social judgment theory. In M. F. Kaplan and S. Schwartz (Eds.), Human judgment and decision processes. New York: Academic Press, 1975.
- Hoepfl, R. T., and Huber, G. P. A study of self-explicated utility models. Behavioral Science, 1970, 15, 408-414.
- Hoerl, A. E. Application of ridge analysis to regression problems. Chemical Engineering Progress, 1962, 58, 54-59.
- Hoerl, A. E., and Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 1970a, 12, 55-67.
- Hoerl, A. E., and Kennard, R. W. Ridge regression: Applications to non-orthogonal problems, Technometrics, 1970b, 12, 69-82.
- Hoffman, P. J. The paramorphic representation of clinical judgment. Psychological Bulletin, 1960, 47, 116-131.
- Huber, G. P., Daneshgar, R., and Ford, D. L. An empirical comparison of five utility models for predicting job preferences. Organizational Behavior and Human Performance, 1971, 6, 267-282.
- Keeney, R. L., and Raiffa, H. Decisions with multiple objectives. New York: Wiley, 1976.
- Keren, G., and Newman, J. R. Additional considerations with regard to multiple regression and equal weighting. Organizational Behavior and Human Performance, 1978, 22, 143-164.



- Kerlinger, F. N., and Pedhazur, E. J. Multiple regression in behavioral research. New York: Holt, Rinehart, & Winston, 1973.
- Klahr, D. Decision making in a complex environment: The use of similarity judgments to predict preferences. Management Science, 1969, 15, 595-618.
- MacCrimmon, K. R. An overview of multiple objective decision making. In J. L. Cochrane and M. Zeleney (Eds.), Multiple criteria decision making. Columbia, S. C.: University of South Carolina Press, 1973.
- Marquardt, D. W., and Snee, R. D. Ridge regression in practice. The American Statistician, 1975, 29, 3-19.
- McClelland, G. Equal versus differential weighting for multiattribute decision: There are no free lunches. Psychological Bulletin, 1979 86 (in press).
- Mettfessel, M. A proposal for quantitative reporting of comparative judgments. Journal of Psychology, 1947, 24, 229-235.
- Mosteller, F., and Tukey, J. W. Data analysis and regression. Reading, Mass.: Addison-Wesley, 1977.
- Newman, J. R. Differential weighting in multi-attribute utility measurement: When it should not and when it does make a difference. Organizational Behavior and Human Performance, 1977, 20, 312-324.
- Nisbett, R. E., and Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. Psychological Review, 1977, 84, 231-259.
- Nystedt, L., and Magnussion, D. Integration of information in a clinical judgment task: An empirical comparison of six models. Perceptual and Motor Skills, 1975, 40, 343-356.
- O'Connor, M. The application of multi-attribute scaling procedures to the development of indices of water quality. Unpublished doctoral dissertation, University of Michigan, 1972.



- Oskamp, S. Clinical judgment from the MMPI: Simple or complex? Journal of Clinical Psychology, 1967, 23, 411-415.
- Permut, S. E. Cue utilization patterns in student-faculty evaluation. Journal of Psychology, 1973, 83, 41-48.
- Price, B. Ridge regression: Application to non-experimental data. Psychological Bulletin, 1977, 84, 759-766.
- Schmitt, N. Comparison of subjective and objective weighting strategies in changing task situations. Organizational Behavior and Human Performance, 1978, 21, 171-188.
- Schmitt, N., Coyle, B. W., and Saari, B. B. Types of task information feedback in multiple-cue probability learning. Organizational Behavior and Human Performance, 1977, 18, 316-328.
- Schmitt, N., and Levine, R. L. Statistical and subjective weights: Some problems and proposals. Organizational Behavior and Human Performance, 1977, 20, 15-39.
- Sherif, M., and Hovland, C. I. Social judgment. New Haven, Conn.: Yale University Press, 1961.
- Slovic, P. Analyzing the expert judge: A descriptive study for stockbroker's decision processes. Journal of Applied Psychology, 1969, 53, 255-263.
- Slovic, P., Fleissner, D., and Bauman, W. S. Analyzing the use of information in investment decision making: A methodological proposal. Journal of Business, 1972, 45, 283-301.
- Slovic, P., and Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgments. Organizational Behavior and Human Performance, 1971, 6, 649-744.
- Stevens, S. S. On the psychophysical law. Psychological Review, 1957, 64, 153-181.

- Summers, D. A., Taliaferro, J. D., and Fletcher, D. J. Subjective weights vs. objective description of judgment policy. Psychonomic Science, 1970, 18, 249-250.
- Torgerson, W. S. Theory and methods of scaling. New York: Wiley, 1958.
- Wainer, H. Estimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 1976, 83, 213-217.
- Wainer, H. On the sensitivity of regression and regressors. Psychological Bulletin, 1978, 85, 267-273.
- Ward, J. E. Comments on "the paramorphic representation of clinical judgment." Psychological Bulletin, 1962, 59, 74-76.
- Wilks, S. S. Weighting systems for linear functions of correlated variables when there is no dependent variable. Psychometrika, 1938, 3, 23-40.
- Winer, B. J. Statistics and data analysis: Trading bias for reduced mean squared error. Annual Review of Psychology, 1978, 29, 647-681.
- Ynetema, D. B., and Torgerson, W. S. Man-computer cooperation in decisions requiring common sense. IRE Transactions of the Professional Group on Human Factors in Electronics, 1961, HFE-2, 20-26.

CONTRACT DISTRIBUTION LIST  
(Unclassified Technical Reports)

Director Advanced Research Projects Agency Attention: Program Management Office 1400 Wilson Boulevard Arlington, Virginia 22209	2 copies
Office of Naval Research Attention: Code 455 800 North Quincy Street Arlington, Virginia 22217	3 copies
Defense Documentation Center Attention: DDC-TC Cameron Station Alexandria, Virginia 22314	12 copies
DCASMA Baltimore Office Attention: Mr. K. Gerasim 300 East Joppa Road Towson, Maryland 21204	1 copy
Director Naval Research Laboratory Attention: Code 2627 Washington, D.C. 20375	6 copies

Revised August 1978

SUPPLEMENTAL DISTRIBUTION LIST  
(Unclassified Technical Reports)

Department of Defense

Director of Net Assessment  
Office of the Secretary of Defense  
Attention: MAJ Robert G. Gough, USAF  
The Pentagon, Room 3A930  
Washington, DC 20301

Assistant Director (Net Technical Assessment)  
Office of the Deputy Director of Defense  
Research and Engineering (Test and  
Evaluation)  
The Pentagon, Room 3C125  
Washington, DC 20301

Director, Defense Advanced Research  
Projects Agency  
1400 Wilson Boulevard  
Arlington, VA 22209

Director, Cybernetics Technology Office  
Defense Advanced Research Projects Agency  
1400 Wilson Boulevard  
Arlington, VA 22209

Director, ARPA Regional Office (Europe)  
Headquarters, U.S. European Command  
APO New York 09128

Director, ARPA Regional Office (Pacific)  
Staff CINCPAC, Box 13  
Camp H. M. Smith, Hawaii 96861

Dr. Don Hirta  
Defense Systems Management School  
Building 202  
Ft. Belvoir, VA 22060

Chairman, Department of Curriculum  
Development  
National War College  
Ft. McNair, 4th and P Streets, SW  
Washington, DC 20319

Defense Intelligence School  
Attention: Professor Douglas E. Hunter  
Washington, DC 20374

Vice Director for Production  
Management Office (Special Actions)  
Defense Intelligence Agency  
Room 1E863, The Pentagon  
Washington, DC 20301

Command and Control Technical Center  
Defense Communications Agency  
Attention: Mr. John D. Hwang  
Washington, DC 20301

Department of the Navy

Office of the Chief of Naval Operations  
(OP-951)  
Washington, DC 20450

Office of Naval Research  
Assistant Chief for Technology (Code 200)  
800 N. Quincy Street  
Arlington, VA 22217

Office of Naval Research (Code 230)  
800 North Quincy Street  
Arlington, VA 22217

Office of Naval Research  
Naval Analysis Programs (Code 431)  
800 North Quincy Street  
Arlington, VA 22217



Office of Naval Research  
Operations Research Programs (Code 434)  
800 North Quincy Street  
Arlington, VA 22217

Office of Naval Research  
Information Systems Program (Code 437)  
800 North Quincy Street  
Arlington, VA 22217

Director, ONR Branch Office  
Attention: Dr. Charles Davis  
536 South Clark Street  
Chicago, IL 60605

Director, ONR Branch Office  
Attention: Dr. J. Lester  
495 Summer Street  
Boston, MA 02210

Director, ONR Branch Office  
Attention: Dr. E. Gloye  
1030 East Green Street  
Pasadena, CA 91106

Director, ONR Branch Office  
Attention: Mr. R. Lawson  
1030 East Green Street  
Pasadena, CA 91106

Office of Naval Research  
Scientific Liaison Group  
Attention: Dr. M. Bertin  
American Embassy - Room A-407  
APO San Francisco 96503

Dr. A. L. Slafkosky  
Scientific Advisor  
Commandant of the Marine Corps (Code RD-1)  
Washington, DC 20380

Headquarters, Naval Material Command  
(Code 0331)  
Attention: Dr. Heber G. Moore  
Washington, DC 20360

Dean of Research Administration  
Naval Postgraduate School  
Attention: Patrick C. Parker  
Monterey, CA 93940

Superintendent  
Naval Postgraduate School  
Attention: R. J. Roland, (Code 52R1)  
C3 Curriculum  
Monterey, CA 93940

Naval Personnel Research and Development  
Center (Code 305)  
Attention: LCDR O'Bar  
San Diego, CA 92152

Navy Personnel Research and Development  
Center  
Manned Systems Design (Code 311)  
Attention: Dr. Fred Muckler  
San Diego, CA 92152

Naval Training Equipment Center  
Human Factors Department (Code N215)  
Orlando, FL 32813

Naval Training Equipment Center  
Training Analysis and Evaluation Group  
(Code N-00T)  
Attention: Dr. Alfred F. Snodde  
Orlando, FL 32813

Director, Center for Advanced Research  
Naval War College  
Attention: Professor C. Lewis  
Newport, RI 02840

Naval Research Laboratory  
Communications Sciences Division (Code 54)  
Attention: Dr. John Shore  
Washington, DC 20375

Dean of the Academic Departments  
U.S. Naval Academy  
Annapolis, MD 21402

Chief, Intelligence Division  
Marine Corps Development Center  
Quantico, VA 22134

#### Department of the Army

Deputy Under Secretary of the Army  
(Operations Research)  
The Pentagon, Room 2E621  
Washington, DC 20310

THIS PAGE IS BEST QUALITY PRACTICABLE  
FROM COPY FURNISHED TO DOD

Director, Army Library  
Army Studies (ASDIRS)  
The Pentagon, Room 1A534  
Washington, DC 20310

U.S. Army Research Institute  
Organizations and Systems Research Laboratory  
Attention: Dr. Edgar M. Johnson  
5001 Eisenhower Avenue  
Alexandria, VA 22333

Director, Organizations and Systems  
Research Laboratory  
U.S. Army Institute for the Behavioral  
and Social Sciences  
5001 Eisenhower Avenue  
Alexandria, VA 22333

Technical Director, U.S. Army Concepts  
Analysis Agency  
8120 Woodmont Avenue  
Bethesda, MD 20014

Director, Strategic Studies Institute  
U.S. Army Combat Developments Command  
Carlisle Barracks, PA 17013

Commandant, Army Logistics Management Center  
Attention: DRXMC-LS-SCAD (ORSA)  
Ft. Lee, VA 23801

Department of Engineering  
United States Military Academy  
Attention: COL A. F. Grum  
West Point, NY 10996

Marine Corps Representative  
U.S. Army War College  
Carlisle Barracks, PA 17013

Chief, Studies and Analysis Office  
Headquarters, Army Training and Doctrine  
Command  
Ft. Monroe, VA 23351

Commander, U.S. Army Research Office  
(Durham)  
Box CM, Duke Station  
Durham, NC 27706

#### Department of the Air Force

Assistant for Requirements Development  
and Acquisition Programs  
Office of the Deputy Chief of Staff for  
Research and Development  
The Pentagon, Room 4C331  
Washington, DC 20330

Air Force Office of Scientific Research  
Life Sciences Directorate  
Building 410, Bolling AFB  
Washington, DC 20332

Commandant, Air University  
Maxwell AFB, AL 36112

Chief, Systems Effectiveness Branch  
Human Engineering Division  
Attention: Dr. Donald A. Topmiller  
Wright-Patterson AFB, OH 45433

Deputy Chief of Staff, Plans, and  
Operations  
Directorate of Concepts (AR/XOCCC)  
Attention: Major R. Linhard  
The Pentagon, Room 4D 1047  
Washington, DC 20330

Director, Advanced Systems Division  
(AFHRL/AS)  
Attention: Dr. Gordon Eckstrand  
Wright-Patterson AFB, OH 45433

Commander, Rome Air Development Center  
Attention: Mr. John Atkinson  
Griffis AFB  
Rome, NY 13440

IRD, Rome Air Development Center  
Attention: Mr. Frederic A. Dion  
Griffis AFB  
Rome, NY 13440

HQS Tactical Air Command  
Attention: LTCOL David Dianich  
Langley AFB, VA 23665

THIS PAGE IS BEST QUALITY FRAGMENTS  
FROM COPY FURNISHED TO DDC

Other Government Agencies

Chief, Strategic Evaluation Center  
Central Intelligence Agency  
Headquarters, Room 2G24  
Washington, DC 20505

Director, Center for the Study of  
Intelligence  
Central Intelligence Agency  
Attention: Mr. Dean Moor  
Washington, DC 20505

Mr. Richard Heuer  
Methods & Forecasting Division  
Office of Regional and Political Analysis  
Central Intelligence Agency  
Washington, DC 20505

Office of Life Sciences  
Headquarters, National Aeronautics and  
Space Administration  
Attention: Dr. Stanley Deutsch  
600 Independence Avenue  
Washington, DC 20546

Other Institutions

Department of Psychology  
The Johns Hopkins University  
Attention: Dr. Alphonse Chapanis  
Charles and 34th Streets  
Baltimore, MD 21218

Institute for Defense Analyses  
Attention: Dr. Jesse Orlansky  
400 Army Navy Drive  
Arlington, VA 22202

Director, Social Science Research Institute  
University of Southern California  
Attention: Dr. Ward Edwards  
Los Angeles, CA 90007

Perceptronics, Incorporated  
Attention: Dr. Amos Freedy  
6271 Variel Avenue  
Woodland Hills, CA 91364

Stanford University  
Attention: Dr. R. A. Howard  
Stanford, CA 94305

Director, Applied Psychology Unit  
Medical Research Council  
Attention: Dr. A. D. Baddeley  
15 Chaucer Road  
Cambridge, CB 2EF  
England

Department of Psychology  
Brunel University  
Attention: Dr. Lawrence D. Phillips  
Uxbridge, Middlesex UB8 3PH  
England

Decision Analysis Group  
Stanford Research Institute  
Attention: Dr. Miley W. Merkhofer  
Menlo Park, CA 94025

Decision Research  
1201 Oak Street  
Eugene, OR 97401

Department of Psychology  
University of Washington  
Attention: Dr. Lee Roy Beach  
Seattle, WA 98195

Department of Electrical and Computer  
Engineering  
University of Michigan  
Attention: Professor Kan Chen  
Ann Arbor, MI 94135

Department of Government and Politics  
University of Maryland  
Attention: Dr. Davis B. Bobrow  
College Park, MD 20747

Department of Psychology  
Hebrew University  
Attention: Dr. Amos Tversky  
Jerusalem, Israel

Dr. Andrew P. Sage  
School of Engineering and Applied  
Science  
University of Virginia  
Charlottesville, VA 22901

THIS PAGE IS BEST QUALITY PRACTICABLE  
FROM COPY FURNISHED TO DDC

Professor Raymond Tanter  
Political Science Department  
The University of Michigan  
Ann Arbor, MI 48109

Professor Howard Raiffa  
Morgan 302  
Harvard Business School  
Harvard University  
Cambridge, MA 02163

Department of Psychology  
University of Oklahoma  
Attention: Dr. Charles Gettys  
455 West Lindsey  
Dale Hall Tower  
Norman, OK 73069

Institute of Behavioral Science #3  
University of Colorado  
Attention: Dr. Kenneth Hammond  
Room 201  
Boulder, Colorado 80309

Decisions and Designs, Incorporated  
Suite 600, 8400 Westpark Drive  
P.O. Box 907  
McLean, Virginia 22101



REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 001922-5-T	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Importance Weight Assessment for Additive, Riskless Preference Functions: A Review.		5. TYPE OF REPORT & PERIOD COVERED Technical 10/77 - 12/78
7. AUTHOR(s) 10 Richard S. / John <del>and</del> Ward / Edwards		6. PERFORMING ORG. REPORT NUMBER - SSRI 78-5
15		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0074 Prime Contract
9. PERFORMING ORGANIZATION NAME AND ADDRESS Social Science Research Institute University of Southern California Los Angeles, California 90007		10. PROGRAM ELEMENT PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd Arlington, VA 22209		12. REPORT DATE December 1978
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Decisions & Desings, Inc. 8400 Westpark Drive, Suite 600 McLean, VA 22101 (under contract from Office of Naval Research)		13. NUMBER OF PAGES 51
15. SECURITY CLASS. (of this report) unclassified		16. DECLASSIFICATION/DOWNGRADING SCHEDULE
17. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited 9 Research rept. Oct 77 - Dec 78		
18. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) 12 72 p.		
13. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) direct subjective estimation      ridge regression indirect holistic estimation      paired comparisons importance weights fractionation		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) One of the more useful tools in decision analysis is the riskless, additive multiattribute utility (MAU) model. The most difficult task in the application of MAU models is that of estimating the importance weight parameters. Two general approaches to the weight estimation problem are extensively reviewed in the present paper: direct subjective estimation and indirect holistic estimation. Various methods for directly assessing importance weights are catalogued, including ranking, fractionation, subjective-estimate methods,		

unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

↙ and paired-comparison procedures, and their relationship to one another is discussed. The so-called indirect holistic methods, including unbiased and biased regression analyses, the ANOVA and fractional ANOVA paradigms, and the indifference techniques of pricing out and trading off to the most important dimension, are all explained with particular emphasis on their common relationship to the general linear model. ↗

A critical review of the literature comparing direct subjective estimation to indirect holistic methods revealed that the conclusions reached by Slovic and Lichtenstein in 1971 are no longer justifiable, if they ever were. Many recent studies are cited in which subjective and statistical weights yield high convergent validity, contrary to the "serious discrepancies between subjective and objective relative weights" focused on in earlier reviews. In addition, several recent studies have established strong evidence of criterion validity for both subjective and statistical weights.

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)